DOCUMENT RESUME

ED 301 120 HE 021 988

| | |
|---|---|
| AUTHOR | Marchese, Theodore J. |
| TITLE | Assessment Update. Third Down, Ten Years to Go. |
| INSTITUTION | American Association for Higher Education, Washington, D.C. |
| PUB DATE | Dec 87 |
| NOTE | 7p. |
| AVAILABLE FROM | American Association for Higher Education, One Dupont Circle, Suite 600, Washington, DC 20036 ($3.00). |
| PUB TYPE | Reports - Descriptive (141) -- Journal Articles (080) |
| JOURNAL CIT | AAHE Bulletin; v40 n4 p3-8 Dec 1987 |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Educational Assessment; *Educational Policy; Educational Quality; Evaluation; Higher Education; *Institutional Evaluation; *Outcomes of Education; *Role of Education; State Programs; *Student Development |
| IDENTIFIERS | *College Outcomes Assessment |

ABSTRACT

Key aspects of assessment's rise as an issue are reported, and several lines of inquiry labeled assessment (and their assumptions) are examined. Concepts of assessment originated decades ago, but until recently people in higher education hearing the word did not view it as they do today. The early 1980s saw the argument that traditional gauges of quality (an institution's resources or reputational ranking) were of less value than a look at its contributions to the student learning. In 1985-86, this view of quality gained currency within the academy and among state policymakers. Colleges were being asked about their contribution to student learning and how they assure it. State interest in postsecondary assessment grew, and in 1987, the tone of state-level discourse about assessment turned more aggressive because there was a sense that colleges were dragging their feet. Nonetheless, a majority of states have not yet mandated anything, although most state leaders remain open to argument. Approaches to assessment are discussed with a focus on the following six assessment traditions: the assessment center; assessment as learning; assessment as program monitoring; assessing student learning and growth; assessment as standardized testing; and the senior examiner. The underlying assumptions in assessment across these six traditions are an eye on performance, a matter of method, and the improvement of performance. (SM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED301120



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

# THIRD DOWN, TEN YEARS TO GO

*by Theodore J. Marchese*

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AAHE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*Important developments occurred in the assessment arena this year. An issue that caused a stir in 1985, and gathered momentum in 1986, surged forward in 1987 under legislative banners of "undergraduate reform," "quality," and "higher performance"—and a renewed sense that as state goals these could be realized through directives to assess.*

*In this article I'll report key aspects of assessment's rise as an issue; examine several lines of inquiry labeled "assessment" and their assumptions; then ask, where are we with this issue?*

Concepts of assessment originated decades ago in corporate and academic circles. But until three years ago, when people in higher education heard the word, most thought of certain research traditions, maybe of Alverno College, and of a few states that had just introduced college-level testing programs.

Debates that began within higher education worked to change that picture. In the early '80s, Alexander Astin took up the argument that traditional gauges of quality—an institution's resources (inputs) or its reputational ranking—were of less value than a look at its contributions to student learning, the "value added" from its teaching or "outcomes" of talent development.

In late '84 and early '85, that view found expression in two influential reports. NIE's *Involvement in Learning* argued that quality was a function of contribution to learning, and that to strengthen learning one needed a) student involvement, b) high expectations, and c) assessment with feedback. AAC's *Integrity in the College Curriculum* also made a learning-assessment linkage, calling it "scandalous" that we fail to assess the impacts of our teaching.

A subtle but important development through 1985-86 was the way this view of "quality" gained currency within the academy. It came to permeate convention speeches, journal articles, faculty debates and, late in 1986, a book by the president of Harvard, Derek Bok's *Higher Learning*.

More significantly, the view took hold among state policymakers. Education-minded governors began asking their colleges, What is your contribution to student learning? Do your students know, and can they do, what your degrees imply? How do you assure that fact? What is the evidence?

A relevant circumstance was that most states had spent the early years of the '80s asking hard questions about school and student performance at the K-12 level. A policy response in all 50 states was to impose statewide standardized tests at various grade levels; they call this testing "assessment," and believe it works. When attentions turned at mid-decade to higher education, some legislators thought they spotted familiar problems (needs to boost student and institutional performance) and looked to a familiar solution, "assessment" as testing.

A combination of ideas, then—that quality was a function of learning, that learning should be demonstrated by assessment, and that tests could leverage improvement—undergirded state interest in postsecondary assessment. As recently as 1985, only a few states had something underway in the name of assessment; by the end of 1986, two-thirds of them did. Influential reports issued from state bodies, notably one from the National Governors' Association (NGA) suggestively titled, *A Time for Results*.

Even so, 1986 was primarily a year of study and debate, and of some restraint among the states, the latter in part because of educator outcries (led in part by AAHE). In that year, six states rejected the idea of a statewide rising-junior exam. A more sensitive model of state leadership for assessment seemed ascendant: Colorado, Connecticut, and Virginia rejected statewide testing and instead set time-lines for public institutions to develop their own approaches tailored to local needs.

This year, 1987, the tone of state-level discourse about assess-

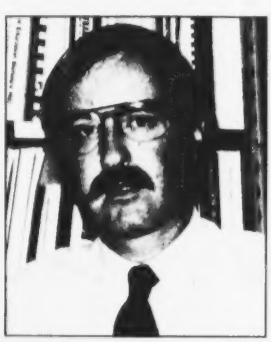
HE 021 988

ment turned more aggressive. A sense developed that colleges were dragging their feet and that it was time to get on with things. In Missouri, Tennessee, Texas, and New Jersey, among other states, significant new directives have come into force; the NGA is organizing states with "mature" programs to coach the followers-behind; in still more states (e.g., California), legislation will come forward in early '88.

Reinforcing this activist mood, Secretary Bennett wrote all governors in July to tell them that educators' lack of accountability stands as one of the great problems of American education, and that they should press forward with assessment. More concretely, the Department has proposed regulations, coming off its authority to recognize accrediting agencies, that would compel all accredited institutions and programs to generate public evidence of student achievement.

To recapitulate, ideas that originated in internal debates about quality were taken up by external parties and translated into mandates to assess. In the course of rapid events, 1985's question—Is this a fad?—has been answered: *No.* Will we see further mandates to assess? *Yes.*

Despite these tides at work, the larger situation remains fluid. A majority of states have not yet mandated anything; existing mandates exhibit no single pattern; some state programs, fashioned with faculty advice, are sensible; and most state leaders remain open to intelligent argument.



*Ted Marchese has been vice president of AAHE since 1982.*

Ideas that originated in internal debates about quality were taken up by external parties and translated into mandates to assess.

While there is time, then, it's important to rehearse higher education's traditions of assessment.

## Approaches to Assessment

Assessment is not something just invented; a rich variety of approaches to knowing about student learning has evolved, through decades of research and campus experience, so that we enter the arena with many options—should we choose to advance them.

The senior declamation and oral examination—public ways of validating the learning of graduates—were common in the 19th century American college. The practice declined in the face of the elective system and credit-hour-driven degree requirements, but not without protest: Abraham Flexner and President Lowell of Harvard in the first years of this century decried the effects, asking how otherwise could we assure the competence of our graduates? Partly as a response, longitudinal studies of college-level learning were attempted in the '20s and '30s (the most famous in Pennsylvania, from 1928-32), and in the '30s many universities established "testing bureaus" to adduce evidence of student achievement.

Since 1940, at least *six* assessment traditions or lines of work can be identified, each with salience today, each starting with its own set of questions. In practice, the methods each evokes are rarely used in isolation; elements from several approaches typically appear in a given college's assessment program. For conceptual purposes only, then, I'll discuss the six separately, then turn to their common elements for a deeper understanding of assessment.

**1. The Assessment Center.** This term describes a process (not a place) aimed originally at problems of personnel selection, driven by the question, how can we develop a larger, more accurate picture of candidate abilities to perform a prospective task?

The approach arose from academic research in the '30s in Germany, England, and the U.S., which argued that human performance was too complex a phenomenon to gauge with paper and pencil tests, that in selecting people for a given role one needed more comprehensive information. Since there are *observable* aspects of performance, the argument continued, as a cornerstone for selection one could create simulations or exercises permitting the direct observation of the intended performance.

In 1940 the British Army, to select candidates for officer training, set up a system that examined candidate backgrounds, used tests and in-depth interviews, but then also contrived realistic performance situations: candidates had to give a morale speech, function in a leaderless patrol, improvise an escape, and so on. Data from *all* these "samples of performance" were pooled, judged by trained experts, and became the basis for selection.

In the U.S., the O.S.S. used these methods to select spies for World War II assignments; subsequent evaluations asserted the superiority of the method over simpler approaches. AT&T set up an assessment center for executive identification in 1954, with longitudinal research again tending to demonstrate the method's effectiveness. Through the '60s and '70s, assessment centers spread to government and over 2,000 corporations; one now finds the device used to help select foreign-service officers, IBM executives, and high school principals.

In higher education today, two parallel developments are the ten-year's work of the American Assembly of Collegiate Schools of Business to create a performance-

based vehicle to assess the knowledge and abilities of business-school students (with data used for program evaluation), and the Carnegie-financed work of Lee Shulman and colleagues at Stanford to create performance-based assessments for the new National Board for Professional Teaching Standards.

Shulman's work captures the essence of the method: it starts with the question, what is it that effective teachers know and can do?, then asks, what combination of exercises will allow us to observe and identify such people?

**2. Assessment as Learning.** Fifteen years ago, educators at Alverno and a few other colleges looked at the assessment-center method and asked not a selection question but a teacher's question: if the components of effective performance are knowable (and learnable), why can't they systematically be taught for? Rather than a one-time judgment at the end, can't the components (and their assessment) be gone at continuously, from day one, in individual classrooms?

Alverno has done this through a curriculum specifying performance levels for eight abilities, with student achievement overseen by faculty teams and external examiners who assess cumulative learning beyond the level of the individual course.

Alverno added a new element to the original center method, *feedback to the learner*, thereby converting a selection tool into one for human development. In each of its classrooms, then, faculty identify particular abilities their course will attend to; convey clear (high) expectations to students about intended performance; arrange class activity so that the requisite knowing and doing come to be exercised, observed, and judged; then provide continuous feedback to students on their performance, with two aims: increments of improvement, and the student's development over time of abilities to *self*-assess.

Alverno provides continuous feedback to students on their performance, with two aims: increments of improvement, and the student's development of abilities to self-assess.

Alverno's curricular plan is distinctive, but in the classroom its faculty members' approach is the ordinary (if seldom realized) prescription for "good teaching" and "active learning": set clear expectations, evoke student involvement, assess and provide feedback. Alverno's "self-assessing" students are every college's ideal of the reflective graduate who continues to learn. What sets Alverno apart is its faculty's restless introspection about forms of academic work, assessment, and feedback that indeed promote learning. (Alverno stands nearly alone, too, in its commitment to long-term evaluation of the impacts of its approaches—which do work, powerfully, and endure in graduates.)

Alverno faculty have introduced concepts of "assessment as learning" to corporate trainers and to faculty in dozens of other settings (notably in medical education). A related (though not connected) line of work is that of K. Patricia Cross, this fall at Miami-Dade (a community-college assessment leader), where efforts are underway to identify course goals and learn how best forms of teacher feedback can aid their attainment by students.

**3. Assessment as Program Monitoring.** In the late '30s and early '40s, Ruth Eckert, C. Robert Pace, and colleagues at Minnesota conducted elaborate research aimed at the question, how effective is our General College in achieving the University's goals for general education? They built their evaluations around program goals; used multiple measures, repeated over time; and chose methods couched to provide feedback for program strengthening. Their unit of analysis was the *program* (a general-education sequence), not the individual student; the purpose was entirely internal and formative and not for external accountability.

The Minnesota approach prompted similar studies in the '40s at Chicago and in the '50s and '60s at Michigan State, work that influenced the development of accreditation, program review, and institutional research, and left behind valuable literature and instruments.

Withal, it's hard to find large-scale, longitudinal studies of this type in more recent years (except at Alverno). Professional evaluators have long since left higher-education studies for fields where their skills were more sought after (evaluating health care, Head Start, etc.). An interesting turn is that Derek Bok has induced Richard Light, president of the American Evaluation Association, to lead an "assessment seminar" at Harvard that is asking classic program-evaluation questions: what is the impact of Harvard's writing courses on student abilities to write?

The fullest application of this line of work today is at the University of Tennessee, Knoxville in a program coordinated by Trudy Banta (a member of AEA). At UT-K, each of the 105 departments and programs gathers yearly multiple forms of evidence about its effectiveness (student achievement, completion, satisfaction, career success, etc.); each department's approach is self-tailored, continuous, and geared to internal improvement. Rather than "evaluation," UT-K's approach is a form of "program monitoring" and quality control; it is "assessment" in the sense of working off stated goals, the continuous use of multiple measures, a premium on feedback, and an emphasis on student learning as the key indicator of quality.

**4. Assessing Student Learning and Growth.** This line of work begins with questions like, what happens to students as a result of their college experience? What changes do we induce over two or four years in their knowledge, thinking skills, values, attitudes, or motivations?

This is a *long* line of work, inaugurated in the '30s, that had an earlier culmination in Feldman and Newcomb's *The Impact of College on Students* (1969), which cited over a thousand studies. Alexander Astin's *Four Critical Years* (1977) was in this tradition. But it has not been an active line of inquiry in recent years, for a variety of reasons. The body of knowledge that remains is partially dated and hard to use: this was research driven by questions of disciplinary importance more than by the questions of teachers or administrators.

A culmination of researchers' interest in student development is the value-added concept, which aptly shifts the focus in quality determinations from program resources to program results, and which poses a good question of a college: what is your contribution to student learning?

For the past three years, a FIPSE-supported consortium of seven colleges and universities coordinated by Astin has attempted to implement value-added concepts. A final report has yet to issue, but participant stories speak to frustrations of instrumentation and method. Hard as it is to do—it is beyond doing, some people argue, and unnecessary—the colleges report that conceptually, at least, value-added is a great way to raise questions—which fundamentally is why one does assessment.

**5. Assessment as Standardized Testing.** There is no single question that drives the use of tests; indeed, it's hard to state a question that somebody hasn't tried to devise a test for. The ubiquity of tests has led some academics who know little about assessment to think the term *means* testing.

Value-added aptly shifts the focus in quality determination from program resources to program results.

Standardized tests are the great invention of American psychology and have been around for decades. In essence a test measures a sample of behavior, under controlled conditions, to raise inferences about performance in some larger domain of behavior. The technology's intent is to gather information in ways that are reliable, useful, and efficient. The variety and possibilities for such tests are wide: they may look to measure achievement, aptitude, values, motivation, satisfaction, and so on; be criterion-referenced or work off norms (or mix the two); be aimed at individual diagnosis or group description; be the product of commercial vendors or of local design; and use methods ranging from the controlled observation of individual performance to machine-scored instruments given to thousands at once.

Given this, few generalizations about the role of testing in post-secondary assessment are possible. In contrast to Europe, where longer-term attention to graduation standards has led to public funding for richer, more complex methods of examining students, in the U.S. the newness of accountability questions and a "quick fix" mentality has tended to promote use of simple, low-cost tests of the paper-and-pencil variety. Especially at the college level, the instruments we have for gauging student learning remain scarce: we do have good entry-level instruments for basic-skills diagnosis, but things get spotty when one turns to outcomes in general education, the major, or to matters of higher-level learning. In professional studies, the situation varies by perceived importance of field: in medical education one finds elaborate methods of testing student competence (including computer-adaptive approaches of the most sophisticated type), but in education one finds the NTE.

Nationally available standardized tests have a significant, limited role to play in a fuller program of institutional assessment. Such tests provide an efficient way to answer at least broader questions about student learning, and allow for normed comparison. But the few outcomes-oriented tests available today often misalign with campus aims and curricula and wind up having little utility for improving anything; several, for example, provide *no* feedback to individual faculty (or students). This picture undoubtedly will change as a new market pushes new test development, but today's circumstance is that, for purposes of accountability or improvement, the larder of tests is pretty bare.

The university with the most experience in using norm-based standardized tests is Northeast Missouri State. Purists may quibble, but that institution has used testing data since the '70s to bring home to its publics that a) it cares about quality, b) it will be accountable, c) its students do learn, and d) performance levels are up.

**6. The Senior Examiner.** This assessment tradition comes from Britain, has 19th century American antecedents, and is today most notably practiced at Swarthmore. Its central questions have been, how can we know and verify the knowledge and abilities of our senior majors? What can we infer from senior performances about a department's program?

The basic approach is that within a discipline—English, for example—faculty from other, comparable English departments will be brought in to examine in depth (orally and in writing) the senior English majors of the host institution. The technique provides a capstone experience for students, and rich feedback to

them on their performance; there is feedback, too, to the department and institution. The approach forces a department to be clear with itself, students, and the visitors about what it is trying to accomplish; relies upon expert, peer judgment; and envisions an iterative process leading to improvement. The procedure provides virtually no information to external parties useful in any accountability sense. This is an "inside" form of professional accountability and quality control, captured by John Harris's aphorism, "Somebody besides the cook should taste the pudding."

Though not labeled "assessment," the common practices of collegiate arts programs—the music major's senior recital, the art major's culminating portfolio juried by practicing artists—offer parallel examples. Two ideas—graduation based upon performance, not just credit-hour accumulation, and performance judgment by parties other than the instructor—underlie the larger spirit.

An extended argument for reviving senior exercises is advanced by Ernest Boyer in his recent book, *College;* his skepticism about test-based outcomes measurement is laced with warnings that quality cannot be achieved "by bypassing the professors and their classrooms." The best, in-depth treatment of the senior-examiner concept may be found in an AAHE Assessment Forum paper written by Bobby Fong. On the campus front, the Association of American Colleges is conducting a FIPSE-supported trial of the concept involving 18 colleges and eleven disciplines, with findings due in 1989.

## The Assessor's Perspective

Across these six traditions one finds underlying assumptions in assessment.

**An Eye on Performance.** Assessors begin with *purposes:* whether the object is to select spies, certify

The ubiquity of tests has led some academics who know little about assessment to think the term *means* testing.

teachers, or evaluate a writing program, clarity about why the assessment will be done, and about the intended "performance," becomes the essential starting point.

In education, the relevant "performance" is *student learning;* the assessor looks to that outcome as the key indicator of quality or effect.

Student learning, in turn, is evidenced not by credit hours or seat time but by *actual abilities to perform,* by what students know and can do, through demonstrations of the use and application of knowledge.

Finally, in education, the assessor wants to judge performance against explicit *criteria* or *standards.* The assessor is not satisfied that a student earned a test score or grade higher than other students; instead the question is, what absolute level of performance is in evidence, and does it meet our standard?

**A Matter of Method.** All forms of assessment, at bottom, rely upon *expert judgment* (even standardized tests). At small points in the chain one may find a technology in use, but running through the larger process—questions of what to assess, how, by whom, and of data use—one finds human judgment. Assessors do not claim objectivity; they seek valid, reliable judgment.

Across the methods, too, one finds judgment by *external parties.* Whether the approach entails outside examiners or a faculty body reviewing achievement across courses, somebody beyond the agent of instruction is asking the questions, what does it add up to? What was learned? Is that good enough?

Because human performance is viewed as complex (as is its judgment), the assessor always wants to employ *multiple methods* (written exams, orals, portfolio analysis, interviews, tests, demonstrations, etc.) in a variety of settings over a period of time as a condition for judgment.

**The Improvement of Performance.** Assessment looks constantly to raise questions and evoke evidence. It is a *formative process*—on-going and integral, not one-time or summative—an engine for self-awareness and change.

The assessor's key tool for promoting student, instructor, program, or institutional improvement is *feedback.* Indeed, more than a tool, feedback becomes an imperative: in choosing a method or instrument, an assessor asks, what quality of information will be realized from it? For whom will that information be useful? Can it enable improvement?

Assessment as a continuous process generating feedback from other parties has a final aim: to develop abilities and habits of *self-assessment.* Applied to students, the goal is to coach them into becoming what Donald Schön calls "reflective practitioners."

## Reflections

The states, as we've seen, have their own tradition of assessment: large-scale testing, road-tested at the K-12 level. Applied to a different enterprise—a college—that tradition can be dangerous, which is why this issue matters . . . and is why it becomes important to know and advance assessment integral to collegiate education. (The failure of academic leaders to know and speak convincingly about assessment is a factor that persuades legislators to press their own formulations.)

Lots of agendas have been pushed into the assessment tent; various parties use the concept to bash faculty, collar students, or advance "the public interest." Others see in it a chance to improve

Assessment *per se* guarantees nothing by way of improvement, no more than a thermometer cures a fever.

teaching and learning, and public support. Clarity about purposes—the why, for whom, to what use—is often a missing element in assessment debates.

The "improvement vs. accountability" dichotomy can be a trap. Collegiate traditions of assessment all imply externality and accountability; even simple forms of mandated testing have as their impulse not single-digit reports but a prompting of results. An institution that responds to a state with a data-gathering effort only, with no eye or connection to improvement, misses the point and sets itself up for a fall.

Assessment *per se* guarantees nothing by way of improvement, no more than a thermometer cures a fever. Only when used in *combination* with good instruction (that evokes involvement, in coherent curricula, etc.), in a *program* of improvement, can the device strengthen education.

Assessment on the cheap won't lead to improvement either, nor are outcomes independent of inputs and resources. The new call for "results" will require investment. From an institution's standpoint, on-going assessment can help make that case.

Rather than a fad, higher education stands three years into a longer-term course of events, a five-to-ten-year stretch across which notions of quality, standards, and accountability will change. Different roads taken in the name of assessment can enrich or reify, enlarge or trivialize one of society's crucial institutions. The road to richer, integral forms of assessment—at the moment, a road less traveled—is the one to look for. ■


**MARK YOUR CALENDARS**

# The Third National Conference on Assessment in Higher Education

**June 8–11, 1988 ■ The Palmer House ■ Chicago, Illinois**

Last June's Assessment Forum conference in Denver served the needs of over two hundred campuses seeking to move ahead with assessment. Participants discovered thoughtful speakers, practical advice, rich resources, and new colleagues. The next step is Chicago, where we invite you to join us June 8–11 at The Palmer House.

Special features of this year's Assessment Forum conference include:

- Special focus on the assessment of general education outcomes, perhaps the most problematic area of assessment and one frequently mandated by state agencies.
- Sessions on a broad range of assessment options, from entry-level testing to integrated senior capstone assessment, to alumni follow up.
- Campus teams, an approach many of you told us was effective last year.
- A pre-conference round of limited-enrollment workshops focusing on questions of implementation.
- Opportunities for post-conference meetings by campus and state delegations.

For further details and registration materials, contact:

THE AAHE ASSESSMENT FORUM

One Dupont Circle, Suite 600
Washington, DC 20036
(202) 293-6440